

# Group Formation for Collaboration in Exploratory Learning using Group Technology Techniques

Mihaela Cocea and George D. Magoulas

London Knowledge Lab, Birkbeck College, University of London  
23-29 Emerald Street, London, WC1N 3QS, UK  
{mihaela, gmagoulas}@dcs.bbk.ac.uk

**Abstract.** Exploratory Learning Environments (ELEs) allow learners to approach a problem in different ways; they are particularly suitable for ill-defined problems where knowledge is less structured and open-ended exploration is allowed. Moreover, multiple solutions which are equally valid are possible and a common and efficient way to convey this is by promoting and supporting students' collaboration. Successful collaboration, however, depends on forming groups in which the activity is relevant for all members of the group. In this paper we present a computational model for group formation for open-ended exploration in ELEs by modelling the various strategies that learners adopt to solve the same task. This is underpinned by Group Technology techniques that use as criteria the learners' strategies and the similarity among them to form groups that match pedagogy considerations. The proposed mechanism is tested in an exploratory learning environment for mathematical generalisation.

**Key words:** collaboration, group formation, exploratory learning environments, Group Technology

## 1 Introduction

Exploratory Learning Environments (ELEs) allow learners to construct and/or explore models by varying their parameters and observing the effects of these variations. ELEs are particularly useful for ill-defined domains [1], where problems are not well structured and there are no clear boundaries between correct and incorrect approaches to solve the task. Moreover, some problems in these domains have several valid solutions but none of them can be considered better than the others. In classroom situations this is often conveyed by discussion between students under the supervision of a teacher/instructor.

Although collaborative learning has been proved successful in classroom situations [2], [3], in computer-supported learning environments it does not seem to lead to the same learning benefits. One of the contributing factors is the way the collaborative groups are formed, as forming efficient groups is very important to ensure an educational benefit from the group interaction [4].

Several approaches have been proposed to address this issue, of which we briefly present only a few. For example, an algorithmic approach has been used in [5] to form homogeneous and heterogeneous groups, and a Genetic Algorithms approach was used to form mixed groups by considering the following

learners' characteristics: learner's personality, competence level, learning style, indicator for collaborative behaviour and indicator for acting as evaluator in peer-assessment. An Ant Colony Optimisation approach is used in [6] to form heterogeneous groups based on personality traits and performance. Fuzzy C-Means clustering for formation of homogeneous and heterogeneous groups is compared to other low complexity algorithms in [7]. A framework for collaborative group formation was proposed in [8] using a constraint satisfaction problem formulation; Semantic Web Technologies and Logic Programming are used. An ontology was proposed as a framework based on learning theories that facilitate group formation and collaborative learning design [9].

Unlike previous work on group formation, our research is in the field of exploratory learning environments and for the domain of mathematical generalisation, where a problem can be solved using different strategies. The goal of collaborative activities is to discuss similarities and differences between strategies adopted by learners to solve the task. Therefore, the criteria used in our approach is the learners' strategies to a particular task and the similarities between those strategies. To address this, we use an approach inspired from Group Technology by defining learners and strategies as property vectors, calculating similarities between them and deriving an incidence matrix on which clustering is performed. This procedure leads to formation of homogeneous groups; however, heterogeneous groups can also be formed by choosing one or more learners from each or some of the homogeneous groups.

We are aware of only one approach based on Group Technology [10] that has been previously used for group formation for educational purposes. This work was developed for distance learning environments with the purpose of grouping learners and learning objects so that students are given the opportunity to learn skills or knowledge that they don't master already [11]. Also, characteristics of solutions have been previously used to support collaboration [12]; more specifically, differences between problem solutions are detected to help students recognize and resolve conflicts between their problem solutions. In contrast to these works, our approach is developed for ill-defined problems and uses similarities rather than differences between approaches to solve the task.

The rest of the paper is structure as follows. The next section describes the learning environment and examples of typical tasks. Section 3 briefly introduces Group Technology (GT) and defines our approach and its relation to GT. Section 4 illustrates the approach using data from a classroom session and Section 5 concludes the paper and presents some directions for future work.

## 2 Mathematical Generalisation with Exploratory Learning

Our research is conducted in the context of *eXpresser* [13], an Exploratory Learning Environment for the domain of mathematical generalisation. The system is intended for 11 to 14 year olds and for classroom use. Individual tasks in *eXpresser* involve building a construction and deriving a rule form it; collaborative

tasks involve discussion on similarities or differences or both between constructions produced individually.

We illustrate here two typical tasks: ‘pond tiling’ and ‘stepping stones’. Pond tiling requires to find a general rule for surrounding any rectangular pond. The construction, several ways of building the construction and their corresponding rules are displayed in Fig. 1; the variables “w” and “h” refer to the width and the height of the pond. The ‘stepping stones’ task requires to build a construction such as the one in Fig. 2(a) and to find a rule for the green (lighter colour) tiles in relation to the red (darker) tiles, i.e. the stepping stones; some construction are expanded for ease of visualisation; the variable “red” refers to the number of red tiles. In these figures, the internal structure of the constructions has been highlighted for clarity. In *eXpresser* all constructions would look the same in the normal course of the task.

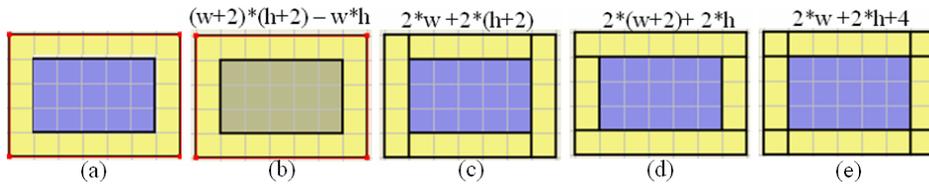


Fig. 1. ‘Pond tiling’ task, constructions and associated rules.

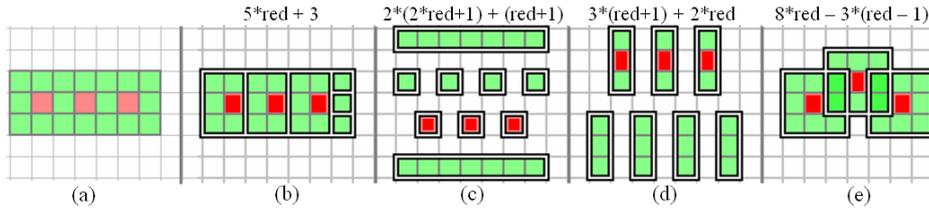


Fig. 2. ‘Stepping Stones’ task, constructions and associated rules.

As illustrated above, each task has multiple solutions, and some of these solutions are similar to each other while others are different. For example, in the ‘pond-tiling’ task the construction in Fig. 1(c) and (e) are similar because they have the same horizontal bars, while the constructions in Fig. 1(d) and (e) are similar for having the same vertical bars. The constructions are illustrated with same pond and the same ‘stepping stones’, respectively; however, learners build constructions of various dimensions. Therefore, in this work the notion of similarity between different strategies refers to structural similarity rather than the exact dimensions of the construction.

As the goal of collaborative activities is to discuss similarities and differences between different learner strategies, the criteria for group formation are: (a) the strategies followed by each learner; (b) the similarities between different strategies. The second criterion is important because students find it difficult to translate between different representations, which in our case correspond to strategies [14]. Therefore, when learners are less experienced, starting with similar strategies facilitates this translation. When learners become more advanced, they can be challenged to translate between totally different strategies. Con-

sequently, we are interested in forming both homogeneous and heterogeneous groups and to this end we use an approach inspired from Groups Technology which is described in the next section.

### 3 Group Formation using Group Technology Techniques

Our approach is inspired by techniques from Group Technology (GT). A brief overview of GT is given below, followed by our approach and the similarities and differences between the two.

Group technology designates a method for group formation of machines and parts in cellular manufacturing systems. The idea is to optimize production of families of parts by creating machine cells. Thus, clusters of machines can be located in close proximity and be responsible for a particular family of parts (and minimize production/transfer time). Therefore, two types of groupings are considered: grouping parts into families and grouping machines into cells. This is referred to as the cell formation problem. Several methods have been used to solve it, such as descriptive procedures, cluster analysis, graph partitioning, artificial intelligence and mathematical programming [10] [15]. We are interested in the clustering analysis approach, which in turn includes several techniques: similarity coefficient, set theoretic, evaluative and other analytic methods [16]. Our approach is using array-based clustering and the similarity coefficient technique.

In array-based clustering, a part-machine incidence matrix is used whose entries are either zero or one. If the entry in row  $i$  and column  $j$  is one, part  $j$  needs to visit machine  $i$ ; if it is zero, no visit is needed. The array-based techniques lead to clusters of parts and machines by rearranging the order of rows and columns to form diagonal blocks of one in the part-machine matrix.

The similarity coefficient techniques involve grouping parts and machines into cells based on similarity measurements between machines and parts, and between machines. These similarity coefficients are used in agglomerative clustering techniques, of which the most popular is single linkage clustering [10].

In our approach, similarity coefficients are used to form the incidence matrix and array-based clustering is then applied to obtain the cells. For us, the parts are the learners, the machines are the strategies of a task, and the goal is to group students according to the strategy (or strategies) they used, taking into account that some strategies are similar to each other while others are different.

In previous work [17] we defined a Case-based Reasoning knowledge representation to delineate the possible strategies that learners could use when solving a task, and an identification mechanism was developed to distinguish which strategy is followed by the learner. Partial and complete solutions are represented as sequences of cases linked by temporal and dependency relations in a Task Model. These are mapped to the learners behaviour in the system using the identification mechanism which is based on similarity metrics for each type of information used in the cases, and the information on the most similar strategy (or strategies) is then stored in a Learner Model.

Consequently, for the purpose of group formation, the strategies are extracted from the Task Model, and the information on what strategies the learners fol-

lowed are extracted from Learner Models. The procedure includes the following phases (see also Fig. 3) which are detailed below:

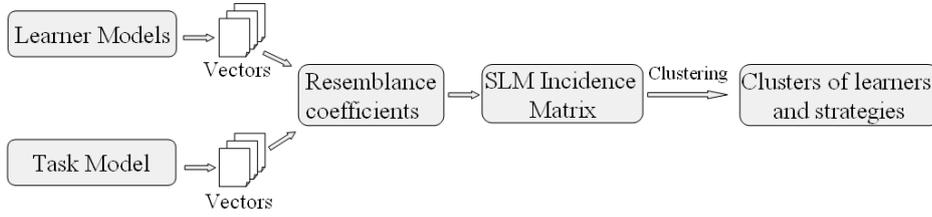
*Phase 1.* Represent all strategies stored in the Task Model as binary vectors that define similarities between them.

*Phase 2.* Retrieve learner strategies from the Learner Models and represent learners as vectors whose elements depict the existence of a relation between a learner' strategy and a strategy stored in the set of task strategies.

*Phase 3.* Define resemblance coefficients and calculate them.

*Phase 4.* Derive the Strategies-Learners Matrix (SLM) from the results of previous step.

*Phase 5.* Perform clustering on SLM.



**Fig. 3.** The procedure for group formation.

**Definition 1.** Let  $S$  be the set of strategies of a task:  $S = \{s_j\}$ ,  $j = 1, 2, \dots, n$ .

Every strategy can be represented as a  $n$ -dimensional vector of 0s and 1s:  $s_j = (s_j^1, s_j^2, \dots, s_j^n)$  where:

$$s_j^i = \begin{cases} 1 & \text{if } s_j \text{ is similar to strategy } s_i \\ 0 & \text{if } s_j \text{ is not similar to strategy } s_i \end{cases}$$

For example, the vectors for the four strategies of the ‘pond-tiling’ task illustrated in Fig. 1 are displayed in Table 1:

**Table 1.** Vectors for the strategies of ‘pond tiling’ task

|     | (b) | (c) | (d) | (e) |
|-----|-----|-----|-----|-----|
| (b) | 1   | 0   | 0   | 0   |
| (c) | 0   | 1   | 0   | 1   |
| (d) | 0   | 0   | 1   | 1   |
| (e) | 0   | 1   | 1   | 1   |

For example, as already mentioned in Section 2, strategy (d) (corresponding to Fig. 1(d)) is similar to itself and to strategy (e); strategy (e) is similar to itself and to strategies (c) and (d). These similarities could be automatically deduced from the existence of structurally similar components like the ones illustrated in Section 2, or they could be formulated by teachers.

**Definition 2.** Let  $L = \{\lambda_k\}$ ,  $k = 1, 2, \dots, m$  be the set of learners.

A learner can be represented as a vector of 0s and 1s:  $\lambda_k = (\lambda_k^1, \lambda_k^2, \dots, \lambda_k^n)$ , where:

$$\lambda_k^i = \begin{cases} 1 & \text{if learner } \lambda_k \text{ used } s_i \text{ strategy} \\ 0 & \text{if learner } \lambda_k \text{ did not use } s_i \text{ strategy} \end{cases}$$

For example, learner A that uses the (b) strategy is represented as (1 0 0 0) and learner B that uses strategy (e) is represented as (0 0 0 1). Sometimes learners use combinations of different strategies; for example, learner C using the (e) and (d) strategies would be represented as (0 0 1 1). This vector formulation is based on the information stored in the Learner Models; thus, learner A and B have in their Learner Models that their constructions are most similar to strategies (b) and (e), respectively, while the Learner Model for learner C indicates that strategies (e) and (d) are most similar.

**Definition 3.** For each learner vector  $\lambda_k$  and each strategy vector  $s_j$ , the following are defined (see also Fig. 4):

1.  $a$  is the number of matching 1s, i.e. the number of strategies contained in both vectors;
2.  $b$  is the number of 1s in  $\lambda_k$  and 0s in  $s_j$ , i.e. the number of strategies followed by the learner which are contained in  $\lambda_k$  but not included in  $s_j$ ;
3.  $c$  is the number of 0s in  $\lambda_k$  and 1s in  $s_j$ , i.e. the number of strategies that the learner did not follow but are included in  $s_j$ .

$$\begin{array}{rcc} \lambda_k: & 0 & \begin{pmatrix} 1 \\ 0 \end{pmatrix} & \begin{pmatrix} 1 \\ 1 \end{pmatrix} & \begin{pmatrix} 0 \\ 1 \end{pmatrix} \\ s_j: & 0 & \begin{pmatrix} 1 \\ 0 \end{pmatrix} & \begin{pmatrix} 1 \\ 1 \end{pmatrix} & \begin{pmatrix} 0 \\ 1 \end{pmatrix} \\ & & b = 1 & a = 1 & c = 1 \end{array}$$

**Fig. 4.** Example for Definition 3.

In line with [11], we use two resemblance coefficients: one for the similarity between learners and strategies, and one for the relevance of each strategy for a particular learner.

**Definition 4.** The similarity coefficient (SC) between a learner  $\lambda_k$  and a strategy  $s_j$  is defined as:  $SC(\lambda_k, s_j) = \frac{a}{a+b+c}$ , for each learner  $\lambda_k \in L$ ,  $k = 1, 2, \dots, m$  and each strategy  $s_j \in S$ ,  $j = 1, 2, \dots, n$ .

This was first defined for use in GT by McAuley [18] and is in fact a Jaccard similarity coefficient, a well known measure of similarity. A study comparing 20 similarity coefficients found it to be the most stable [19].

**Definition 5.** The Relevance Coefficient (RC) of a strategy  $s_j$  for learner  $\lambda_k$  is defined as:  $RC(\lambda_k, s_j) = \frac{a}{a+b}$ , for each learner  $\lambda_k \in L$ ,  $k = 1, 2, \dots, m$  and each strategy  $s_j \in S$ ,  $j = 1, 2, \dots, n$ .

The strategies-learners matrix is defined as:  $SLM = \{c_{ij}\}$ ,  $i \in [1, n]$ ,  $j \in [1, m]$ ,

$$c_{ij} = \begin{cases} 1 & \text{if } RC \geq \theta^{RC} \text{ and } SC \geq \theta^{SC} \\ 0 & \text{otherwise} \end{cases}$$

where  $\theta^{RC}, \theta^{SC} \in (0, 1]$ .

A minimum density of the matrix is necessary to obtain meaningful results. More specifically, each column should have at least a ‘1’, i.e. each learner should follow at least one strategy. Therefore, the minimum density is the number of learners:  $m$ . Consequently, to fulfill the matrix density constraint, the values of  $\theta^{RC}$  and  $\theta^{SC}$  could be defined dynamically for each class. To avoid unnecessary computation, however, the following were established: (a) as the relevance of a

strategy for a learner should play an important role, the value of  $\theta^{RC}$  should not be lower than 0.5; (b) calculate values dynamically only if the density constraint is not satisfied using the value of 0.5 for both thresholds. Therefore, the grouping starts with the value of 0.5 for both thresholds and if the matrix density constraint is not satisfied, the value of  $\theta^{SC}$  is gradually decreased until the constraint is satisfied.

To illustrate the next phase of the procedure, i.e. the clustering, let us consider the matrix displayed in Step 1 of Fig. 5. Rank Order Clustering (ROC), one the most frequently used methods in array-based clustering [15], is used, which involves organising columns and rows in the order of decreasing binary weights. The following procedure is applied which is illustrated in Fig. 5:

*Step 1.* Assign value  $2^{m-j}$  to column  $j$ . Evaluate each row ( $Row_i = \sum_{j=1}^m c_{ij} * 2^{m-j}$ ) and order rows in decreasing order. If there is no change compared to previous order, stop. Else, go to step 2.

*Step 2.* Assign value  $2^{n-i}$  to row  $i$ . Evaluate each column ( $Column_j = \sum_{i=1}^n c_{ij} * 2^{n-i}$ ) and order columns in decreasing order. If there is no change compared to previous order, stop. Else, go to step 1.

| Step 1           |          |    |   |   |   |   | Step 2     |          |    |    |   |   |    | Repeat Step 1 |                  |    |    |   |   |   |   |   |    |
|------------------|----------|----|---|---|---|---|------------|----------|----|----|---|---|----|---------------|------------------|----|----|---|---|---|---|---|----|
| Strategies       | Learners |    |   |   |   |   | Strategies | Learners |    |    |   |   |    | Strategies    | Learners         |    |    |   |   |   |   |   |    |
|                  | 1        | 2  | 3 | 4 | 5 | 6 |            | 1        | 2  | 3  | 4 | 5 | 6  |               | 2 <sup>n-i</sup> | 1  | 3  | 6 | 2 | 5 | 4 |   |    |
| 1                | 1        | 1  | 1 | 0 | 0 | 1 | 57         | 1        | 1  | 1  | 1 | 0 | 0  | 1             | 8                | 1  | 1  | 1 | 1 | 1 | 0 | 0 | 60 |
| 2                | 0        | 1  | 0 | 0 | 1 | 0 | 18         | 3        | 1  | 0  | 1 | 0 | 0  | 1             | 4                | 3  | 1  | 1 | 1 | 0 | 0 | 0 | 56 |
| 3                | 1        | 0  | 1 | 0 | 0 | 1 | 41         | 2        | 0  | 1  | 0 | 0 | 1  | 0             | 2                | 2  | 0  | 0 | 0 | 1 | 1 | 0 | 6  |
| 4                | 0        | 0  | 0 | 1 | 0 | 0 | 4          | 4        | 0  | 0  | 0 | 1 | 0  | 0             | 1                | 4  | 0  | 0 | 0 | 0 | 0 | 1 | 1  |
| 2 <sup>m-j</sup> | 32       | 16 | 8 | 4 | 2 | 1 |            | 12       | 10 | 12 | 1 | 2 | 12 |               | 2 <sup>m-j</sup> | 32 | 16 | 8 | 4 | 2 | 1 |   |    |

Fig. 5. Steps of Rank Order Clustering example.

In this example, the following clusters were formed: learners 1, 3 and 6 with strategies 1 and 3; learners 2 and 5 with strategy 2, and learner 6 with strategy 4. In this particular example the blocks of 1s are clear cut; however, that is rarely the case, showing that clusters are not independent. Also, one strategy may be used by many learners, forming a big cluster; this is known as the “bottleneck machine problem”, when a large number of components need to be processed by one machine. In the context of forming groups for collaboration though these constraints are not considered critical limitations for the formulation of strategies-learners clusters: if clusters are not independent, it means that some learners are using other strategies besides the ones of that cluster; if many learners are using the same strategy, forming a large cluster, it can be broken down in several subgroups for the purpose of the collaborative task.

## 4 An illustrative example

We illustrate the approach presented in the previous section using data from a classroom session where 18 students used *eXpresser* to solve the ‘stepping stones’ task. Out of the 18 learners, 6 used the so-called ‘C’ strategy (Fig. 2(b)) (C), 4 used the ‘HParallel’ strategy (Fig. 2(c)) (H), 2 used the ‘VParallel’ strategy (Fig. 2(d)) (V), 1 used the ‘Squares’ strategy (Fig. 2(d)) (S), 1 used a combination of ‘HParallel’ and ‘VParallel’ strategies (H&V) [17] and the remaining 4

students were either off-task or used non-systematic approaches such as building the construction using individual tiles - see Table 2.

**Table 2.** Distribution of strategies used by learners.

| Strategies         | C | H | V | S | H&V | Other |
|--------------------|---|---|---|---|-----|-------|
| Number of Learners | 6 | 4 | 2 | 1 | 1   | 4     |

A subset of the vectors for strategies and learners is displayed in Table 3. For learners that used the same strategy, only one example is provided; for example, learners  $\lambda_1$  to  $\lambda_6$  have the same vectors and thus only learner  $\lambda_1$  is displayed. The learners that did not follow a systematic approach are excluded. As shown in Table 2, 4 learners used non-systematic approaches to solve the task denoted by ‘Other’. These could also be represented by a distinctive vector which will result in a cluster formed by these learners; however, they are already classified as a distinctive group and, therefore, including them in the grouping mechanism will only lead to unnecessary computations.

**Table 3.** Strategies and learners vectors.

| Strategies | C | H | V | S | Learners | $\lambda_1$ | $\lambda_7$ | $\lambda_{11}$ | $\lambda_{13}$ | $\lambda_{14}$ |
|------------|---|---|---|---|----------|-------------|-------------|----------------|----------------|----------------|
| C          | 1 | 0 | 0 | 1 | C        | 1           | 0           | 0              | 0              | 0              |
| H          | 0 | 1 | 0 | 0 | H        | 0           | 1           | 0              | 0              | 1              |
| V          | 0 | 0 | 1 | 0 | V        | 0           | 0           | 1              | 0              | 1              |
| S          | 1 | 0 | 0 | 1 | S        | 0           | 0           | 0              | 1              | 0              |

Table 4 displays the values of  $RC$  and  $SC$  for each strategy and learner. Using  $\theta^{RC} = 0.5$  and  $\theta^{SC} = 0.5$ , the initial matrix in Table 5 is obtained; applying ROC to it leads to the final matrix in Table 5 and to the following groups:

- (1) Group 1 includes learners  $\lambda_i, i = 1, 2, \dots, 6$  and  $\lambda_{13}$  that adopted the ‘C’ and ‘Squares’ strategies;
- (2) Group 2 includes learners  $\lambda_i, i = 7, 8, \dots, 10$  and  $\lambda_{14}$  that adopted the ‘HParallel’ strategy;
- (3) Group 3 includes learners  $\lambda_i, i = 11, 12$  that adopted the ‘VParallel’ strategy.

The advantages of using this method, as opposed to clustering based only on the strategies used, is that the similarities between different strategies could be modified by the teacher. They could vary from being very strict (a strategy is similar only to itself) to being very relaxed (a strategy is similar to other strategies when there is at least one part that is similar).

For example, a teacher may consider the ‘Squares’ strategy to be similar to the ‘VParallel’ rather than the ‘C’ strategy. Consequently, the strategies vectors would be: (a) ‘C’ strategy (1 0 0 0); (b) ‘HParallel’ strategy (0 1 0 0); (c) ‘VParallel’ strategy (0 0 1 1); (d) ‘Squares’ strategy (0 0 1 1). Using these vectors a new SLM matrix is obtained and the clustering procedure outputs the following groups:

- (1) Group 1 includes learners  $\lambda_i, i = 1, 2, \dots, 6$  that used the ‘C’ strategy;
- (2) Group 2 includes learners  $\lambda_i, i = 7, 8, \dots, 10$  and  $\lambda_{14}$  that used the ‘HParallel’ strategy;
- (3) Group 3 includes learners  $\lambda_i, i = 11, 12$  and  $\lambda_{14}$  that used the ‘VParallel’ and ‘Squares’ strategies.

**Table 4.** Values of Relevance Coefficients (RC) and Similarity Coefficients (SC).

| Strategies |    | $\lambda_1$ | $\lambda_2$ | $\lambda_3$ | $\lambda_4$ | $\lambda_5$ | $\lambda_6$ | $\lambda_7$ | $\lambda_8$ | $\lambda_9$ | $\lambda_{10}$ | $\lambda_{11}$ | $\lambda_{12}$ | $\lambda_{13}$ | $\lambda_{14}$ |
|------------|----|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|----------------|----------------|----------------|----------------|----------------|
| C forward  | RC | 1           | 1           | 1           | 1           | 1           | 1           | 0           | 0           | 0           | 0              | 0              | 0              | 1              | 0              |
|            | SC | 0.5         | 0.5         | 0.5         | 0.5         | 0.5         | 0.5         | 0           | 0           | 0           | 0              | 0              | 0              | 0.5            | 0              |
| HParallel  | RC | 0           | 0           | 0           | 0           | 0           | 0           | 1           | 1           | 1           | 1              | 0              | 0              | 0              | 0.5            |
|            | SC | 0           | 0           | 0           | 0           | 0           | 0           | 1           | 1           | 1           | 1              | 0              | 0              | 0              | 0.5            |
| VParallel  | RC | 0           | 0           | 0           | 0           | 0           | 0           | 0           | 0           | 0           | 0              | 1              | 1              | 0              | 0.5            |
|            | SC | 0           | 0           | 0           | 0           | 0           | 0           | 0           | 0           | 0           | 0              | 1              | 1              | 0              | 0.5            |
| Squares    | RC | 1           | 1           | 1           | 1           | 1           | 1           | 0           | 0           | 0           | 0              | 0              | 0              | 1              | 0              |
|            | SC | 0.5         | 0.5         | 0.5         | 0.5         | 0.5         | 0.5         | 0           | 0           | 0           | 0              | 0              | 0              | 0.5            | 0              |

**Table 5.** Initial and final matrix.

| Initial matrix |          |   |   |   |   |   |   |   |   |    |    |    |    | Final matrix (after ROC) |          |          |          |          |          |          |          |          |          |          |          |          |          |          |          |          |
|----------------|----------|---|---|---|---|---|---|---|---|----|----|----|----|--------------------------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| Strategies     | Learners |   |   |   |   |   |   |   |   |    |    |    |    | Strategies               | Learners |          |          |          |          |          |          |          |          |          |          |          |          |          |          |          |
|                | 1        | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |                          | 14       | 1        | 2        | 3        | 4        | 5        | 6        | 13       | 14       | 7        | 8        | 9        | 10       | 11       | 12       |          |
| C              | 1        | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0  | 0  | 0  | 1  | 0                        | C        | <b>1</b> | 0        | 0        | 0        | 0        | 0        | 0        | 0        |
| H              | 0        | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1  | 0  | 0  | 0  | 1                        | S        | <b>1</b> | 0        | 0        | 0        | 0        | 0        | 0        | 0        |
| V              | 0        | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0  | 1  | 1  | 0  | 1                        | H        | 0        | 0        | 0        | 0        | 0        | 0        | 0        | 0        | <b>1</b> | <b>1</b> | <b>1</b> | <b>1</b> | <b>1</b> | 0        | 0        |
| S              | 1        | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0  | 0  | 0  | 1  | 0                        | V        | 0        | 0        | 0        | 0        | 0        | 0        | 0        | 0        | 1        | 0        | 0        | 0        | 0        | <b>1</b> | <b>1</b> |

The mechanism we developed provides the teachers with groups based on the strategies followed by learners, i.e. the clusters formed as explained above. Using this information, teachers decide the size of groups and how the learners are distributed. Currently, our approach does not include social, cultural or personality factors, which are handled by the teacher. Future work, however, will look at integrating these factors and at automating the group formation in a flexible manner that, for example, will allow teachers to enter constraints such as ‘learner X should never be grouped with learner Y’.

## 5 Conclusions and Future Work

In this paper we presented an approach for group formation inspired from Group Technology. The criteria used in the grouping process are the strategies followed by learners in solving a task and the similarities between different strategies of a particular task. Resemblance coefficients are used to define: (a) the similarity between learners and strategies and (b) the relevance of each strategy for a particular learner. The approach outputs homogeneous groups; however, heterogeneous groups can be formed by choosing one or more learners from each or some of the homogeneous groups.

The model could be refined to better reflect the similarity of the learner’s construction to the strategies in the Task Model. At the moment, when a learner uses more than one strategy, these are all contained in the learner’s vector and are represented by a value of 1. In addition, no special weights are used as strategies are all considered equally important. However, the learner model includes information about the similarity of learner’s construction with each strategy and these could be proportionally reflected in the learner’s vector. Also, the similarities between strategies could be refined to reflect partial similarity rather than having a value of ‘1’ for a range of similarities going from similarity of only one part to full similarity. This would possibly require a modification of the resemblance coefficients and is part of our future work.

## References

1. Lynch, C., Ashley, K., Alevan, V., Pinkwart, N.: Defining ill-defined domains; a literature survey. In: Proc. of the Workshop on Intelligent Tutoring Systems for Ill-Defined Domains at the Intelligent Tutoring Systems Conference. (2006) 1–10
2. Brown, A., Palincsar, A.: Knowing, learning, and instruction. In: Guided, cooperative learning and individual knowledge acquisition. Hillsdale, NJ: Lawrence Erlbaum Associates (1989) 307–336
3. Slavin, R.E.: When and Why Does Cooperative Learning Increase Achievement? In: The RoutledgeFalmer Reader in Psychology of Education. Volume 1 of RoutledgeFalmer Readers in Education. RoutledgeFalmer (2003) 271 – 293
4. Daradoumis, T., Guitert, M., Giménez, F., Marquès, J.M., Lloret, T.: Supporting the composition of effective virtual groups for collaborative learning. In: Proc. of ICCE'02, IEEE Computer Society (2002) 332 – 336
5. Gogoulou, A., Gouli, E., Boas, G., Liakou, E., Grigoriadou, M.: Forming homogeneous, heterogeneous and mixed groups of learners. In: Proc. of PING Workshop. (2007) 33 – 40
6. Graf, S., Bekele, R.: Forming heterogeneous groups for intelligent collaborative learning systems with ant colony optimization. In: ITS'06 Proc. (2006) 217–226
7. Christodouloupoulos, C.E., Papanikolaou, K.: Investigation of group formation using low complexity algorithms. In: Proc. of PING Workshop. (2007) 57 – 60
8. Ounnas, A., Davis, H.C., Millard, D.E.: A framework for semantic group formation in education. *Educ. Tech. Soc.* **12**(4) (2009) 4355
9. Isotani, S., Inaba, A., Ikeda, M., Mizoguchi, R.: An ontology engineering approach to the realization of theory-driven group formation. *I. J. Computer-Supported Collaborative Learning* **4**(4) (2009) 445–478
10. Selim, H.M., Askin, R.G., Vakharia, A.J.: Cell formation in group technology: Review, evaluation and directions for future research. *Comput. Ind. Eng.* **34**(1) (1998) 3 – 20
11. Pollalis, Y.A., Mavrommatis, G.: Using similarity measures for collaborating groups formation: A model for distance learning environments. *Eur. J. Oper. Res.* **193**(2) (2009) 626 – 636
12. de los Angeles Constantino-Gonzalez, M., Suthers, D.D., de los Santos, J.G.E.: Coaching web-based collaborative learning based on problem solution differences and participation. *Int. J. Artif. Intell. Ed.* **13**(2-4) (2003) 263–299
13. Pearce, D., Mavrikis, M., Geraniou, E., Gutierrez, S.: Issues in the design of an environment to support the learning of mathematical generalisation. In: Proc. of EC-TEL'08. (2008) 326–337
14. Ainsworth, S.: The functions of multiple representations. *Comput. Educ.* **33**(2-3) (1999) 131–152
15. Joines, J.A., King, R.E., Culbreth, C.T.: A comprehensive review of production-oriented manufacturing cell formation techniques. *Int. J. Flex. Autom. Integr. Manuf.* **3**(3-4) (1996) 225–265
16. King, J., Nakornchai, V.: Machine-component group formation in group technology: Review and extension. *Int. J. Product. Res.* **20**(2) (1982) 117 – 133
17. Cocea, M., Magoulas, G.: Task-oriented modeling of learner behaviour in exploratory learning for mathematical generalisation. In: Proc. of the 2nd ISEE workshop. (2009) 16–24
18. McAuley, J.: Machine grouping for efficient production. *Production Engineer* **51**(2) (1972) 53 – 57
19. Yin, Y., Yasuda, K.: Similarity coefficient methods applied to the cell formation problem: a comparative investigation. *Comput. Ind. Eng.* **48**(3) (2005) 471 – 489